UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

February 28, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-12*

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: William R. Bell William R. Bell
Senior Mathematical Statistician for Small Area Estimation

Subject: Accuracy and Coverage Evaluation: Correlation Bias

The attached document was prepared. per your request, to assist the Executive Steering Committee on A.C.E. Policy in assessing the data with and without statistical

This report focuses on investigation and estimation of correlation bias in A.C.E. dual system estimates (DSEs) by comparing them to results from demographic analysis (DA). Specifically, I estimate correlation bias for adult male DSEs assuming no correlation bias for adult female DSEs and controlling the male estimates to DA sex ratios.

DSSD Census 2000 Procedures and Operations Memorandum Series B-12*
February 28, 2001

Accuracy and Coverage Evaluation: Correlation Bias

William Bell
U.S. Census Bureau

Table of Contents

Tables

Accuracy and Coverage Evaluation 2000: Correlation Bias

prepared by William Bell

Executive Summary

We compared 2000 A.C.E. estimates aggregated to the national level for age-race-sex groups (Black versus Nonblack race) against results from demographic analysis (DA) to investigate possible correlation bias in the A.C.E. dual system estimates (DSEs). We made these comparisons for both population totals and sex ratios. We then used previously developed methods (Bell 1993) to estimate correlation bias in the DSEs for adult males using the DA sex ratios, assuming no correlation bias for adult females.

What is correlation bias?

Correlation bias can be defined as the error that would result from DSEs based on "perfect data," that is, DSEs based on data with no sampling error and no other biases. Correlation bias results from failure of the general independence assumption underlying DSEs due to

- **causal dependence –** the act of being included in the Census makes someone more likely or less likely to be included in the A.C.E., or

- **heterogeneity** – Census and A.C.E. inclusion probabilities vary over persons within post-strata.

If heterogeneity exists in the sense that those more likely to be missed in the Census are also more likely to be missed in A.C.E., then correlation bias is negative implying underestimation by the DSEs. While causal dependence can lead to either positive or negative biases in DSEs, generally the concern about correlation bias is heterogeneity leading to underestimation.

What evidence do we have of correlation bias in DSEs from A.C.E.?

Historical evidence of correlation bias in DSEs comes from comparisons of results aggregated to the national level against DA estimates for age-sex-race groups (Black versus Nonblack race). Comparisons against DA population totals provide a crude check because of the relatively large uncertainty about the effects on DA totals of errors in estimates of undocumented immigration and emigration. We believe DA sex ratios (number of males over number of females) are more accurate than DA totals and provide a more refined check of correlation bias for adult males (assuming negligible correlation bias for adult females.) Comparisons of 1990 PES estimates (357 post-strata) against 1990 DA estimates (totals and sex ratios) gave evidence of significant correlation bias in DSEs for adult Black males and possible correlation bias for adult Nonblack males, but did not give evidence of correlation bias for children or adult females.

Comparison of 2000 A.C.E. estimates against DA totals and sex ratios reveals the following:

- 2000 DA totals are *lower* than A.C.E. totals for many of the age-race-sex groups. The difference is particularly large for Nonblack males and females 18-29 (−8.4 percent and −5.8 percent, respectively.) These results are hard to explain in terms of "correlation bias." Comparisons of sex ratios for Nonblacks 18-29 are suspect given these results.

- 2000 DA sex ratios for adult Blacks significantly exceed those for A.C.E., strongly suggesting correlation bias in DSEs for adult Black males. DA sex ratios for Nonblacks 30-49 and 50 and over only slightly exceed those from A.C.E. suggesting at most small amounts of correlation bias.

How can we estimate correlation bias in DSEs (for adult males)?

We used various models to estimate correlation bias in adult male DSEs from A.C.E. These models assume no correlation bias for adult females and that the DA sex ratios are accurate. The models produce alternative estimates for the male post-strata that are constrained to aggregate to control totals that reproduce the DA sex ratios for the age-race group. The percent differences between these alternative estimates and the usual DSEs then estimate relative correlation bias for the individual male post-strata. We can think of this approach as estimating the correlation bias at the national level from the comparison of A.C.E. results against DA, and then allocating the aggregate bias to the post-strata in various ways according to the different model assumptions.

This approach produced the following results for the 2000 A.C.E. estimates:

- We estimated significant correlation bias for adult Black males. For the simplest model used (which assumes constant relative correlation bias across post-strata), these estimates for 18-29 and 30-49 (−7.4 and −8.1 percent) are very similar to those in 1990.

- The corresponding estimate of correlation bias for Black males 50 and over is −4.7 percent, compared to −8.2 percent in 1990. The difference may come from significant revisions made to the DA estimates for the cohort of Black males age 65-75 in 2000.

- Corresponding estimates of correlation bias for Nonblack males 30-49 and 50 and over are −.5 and −.7 percent in 2000, which are small though not very different from in 1990.

- Due to the large inconsistencies between DA and A.C.E. estimates for Nonblacks 18-29, both for totals and sex ratios, we concluded that we cannot use the DA sex ratios for estimating correlation bias for Nonblack males 18-29. We proceeded with the assumption of no correlation bias for Nonblack males 18-29 on the grounds that

    - 2000 estimates of relative correlation bias for older Nonblack males are small, and
    - the 1990 estimate of relative correlation bias for Nonblack males 18-29 was small.

Introduction

Dual system estimates (DSEs) are said to contain bias if they systematically underestimate or overestimate the true population. Biases in the sample estimates of the components of the DSE formula can lead to biases in the DSEs, as discussed in Mulry (1991) and Mulry and Spencer (1991,1993). Even in the absence of any of these biases, DSEs can still be subject to another form of bias called *correlation bias*, resulting from failure of a general independence assumption that underlies the DSEs. This independence assumption can fail due either to *causal dependence* between the act of inclusion in the Census and the act of inclusion in the A.C.E., or due to *heterogeneity* across persons in the probabilities of being included in the Census and in the A.C.E. DSEs are constructed within post-strata to reduce heterogeneity in the inclusion probabilities, so that heterogeneity leading to correlation bias exists only if the inclusion probabilities vary across persons *within* a post-stratum. For further general discussion of correlation bias, see Griffin (2000).

When heterogeneity exists it is generally suspected to be of the form where persons (within a post-stratum) more likely to be missed in the Census are also more likely to be missed in the coverage survey (A.C.E.). Correlation bias resulting from this form of heterogeneity is negative, reflecting systematic underestimation of true population by the DSEs. The direction of the effect of causal dependence, if it exists, is less certain. It could be that persons included in the Census are made more aware of the Census process, and hence are more likely to be included in the A.C.E. than those missed by the Census. This type of dependence would lead to underestimation by the DSEs. Or it could be that persons included in the Census feel they have already responded to the Census Bureau, and so are more resistant to being included in the A.C.E. than those missed by the Census. This type of dependence would lead to overestimation by the DSEs. It is hard to say which of these two types of causal dependence might be more likely, thus, it is difficult to say whether correlation bias due to causal dependence would tend to lead to underestimation or overestimation by the DSEs.

Two approaches that have been used to investigate possible correlation bias in DSEs from a given post-stratification are:

1. compare to results from demographic analysis,

2. compare to results from alternative post-stratification DSEs or more general models (e.g., logistic regression).

Both these approaches have strengths and limitations. Two other approaches that have been used that provide some evidence about possible correlation bias in DSEs are triple system estimation (Zaslavsky and Wolfgang 1993) and ethnographic observation. However, both these approaches have been used only in limited test sites, and so were not feasible approaches for assessing correlation bias in estimates from the 2000 A.C.E.

Demographic analysis (DA), discussed by Robinson et al. (1993), has the advantage that its estimates are constructed from administrative data sources some of which (birth and death registration data) are generally believed to be quite accurate. Comparison of DSE results against DA estimates provides an aggregate check for correlation bias whether due to causal dependence or heterogeneity (with some qualifications regarding allowance for other biases, as noted below). However, DA estimates are adversely affected by errors in the administrative data, such as uncertainty about the level of emigration from the U.S., and uncertainty about the level of undocumented immigration. For this reason, DA population estimates (DA totals) are thought to be less accurate than DA *sex ratios* (number of males over number of females). This reflects an assumption that errors in migration estimates are not grossly different for males than for females.

In addition to errors in its administrative data sources, the primary limitation on DA results is a lack of detail. Difficulties in using administrative data to construct estimates of subnational migration mean that subnational DA estimates, while providing useful indicators, are of significantly lesser accuracy than DA national estimates. Also, limited racial detail in the administrative data sources, along with differences in racial classification from the Census, limits separate DA estimates by race to simply Black and Nonblack. This limitation is somewhat more pronounced in 2000 than in 1990 because the allowance of multiple race responses to the 2000 Census creates some uncertainty about appropriate definitions of the Black and Nonblack groups for comparability of DA and A.C.E. results. This is discussed in the next section. These effects are more important for making comparisons to DA totals than to DA sex ratios.

Because of the limitations of DA, to investigate possible correlation bias in the 2000 A.C.E. DSEs we only use DA data at the national level broken down by age (four A.C.E. age groups), race (Black and Nonblack groups), and sex. We use DA totals to provide rather crude checks on possible bias separately for males and females, and also for all children 0-17. We use DA sex ratios to explore in more detail possible correlation bias for adult males assuming no correlation bias for adult females.

The approach of comparing DSEs from a given post-stratification against DSEs from alternative post-stratifications or more general models was used in the research leading to the development of the A.C.E. post-stratification (Griffin 1999, Schindler 2000). One objective in this work was to use comparisons of post-stratified DSEs to estimates from a logistic regression "target" model to develop bias estimates for the DSEs. We found this approach to estimating bias was not very successful, however, probably due to failure of the underlying assumption that the target model produced estimates that were unbiased, or at least substantially less biased than the DSEs considered. More explicitly, the difference between two DSEs will reflect both the difference in their biases and some amount of random error, and unless the bias difference is substantial relative to the amount (standard deviation) of the random error, the resulting estimates of the bias difference will be poor. This was the case in the post-stratification research, where comparisons of DSEs from reasonable post-stratifications against estimates from the target model often resulted in negative estimates of squared bias (Schindler 2000). Thus, while logistic regression models provide an attractive way to investigate possible heterogeneity unaccounted for by the post-stratification, it is difficult to translate these results into correlation bias estimates, per se. This approach to investigating heterogeneity and its consequences will be pursued in further research beyond the 2000 Census, when there is more time to carefully develop and assess the results, but we have not used this approach for estimating correlation bias in the 2000 A.C.E. estimates.

The next section makes comparisons of DA and A.C.E. results to investigate possible correlation bias in the A.C.E. estimates. The section after that uses DA sex ratios and A.C.E. results to develop estimates of correlation bias for adult males assuming no correlation bias for adult females. A final section then discusses a point raised by Wachter and Freedman (1999), namely that estimates of correlation bias using results from DA can be affected by other biases in the A.C.E. data.

Comparing 2000 A.C.E. and DA Estimates

Adjusting the DA estimates for comparability with A.C.E. estimates

For comparison to the 2000 A.C.E. results, we use here DA estimates revised as of February 16, 2001. Definitional differences between DA and A.C.E. require adjustments to the data to make the two sets of estimates comparable. Since we wish to assess possible correlation bias in the A.C.E. estimates, we make adjustments to the DA estimates, and not the other way around. We made the following adjustments to the DA totals for Blacks and Nonblacks to make them comparable to the A.C.E. results. We then computed DA sex ratios from these adjusted totals.

- We subtracted the Census count of the group quarters population from the 2000 DA totals, since the group quarters population is not part of the A.C.E. universe. For 1990 a smaller population was subtracted out as not part of the PES universe.

- We subtracted estimates of Black Hispanics from the DA totals for Blacks and added these estimates to the DA totals for Nonblacks. This was also done for the group quarters population subtracted out as just noted. We need this adjustment because A.C.E. assigns Black Hispanics to its Hispanic race domain (domain 3), not its Black race domain (domain 4). The 1990 PES assigned Black Hispanics to the Black post-strata, so this issue did not arise. The implied DA estimates of Black Hispanics for 2000 were obtained by inflating the Census counts of Black Hispanics by adjustment factors corresponding to the DA estimates of Black undercount, since separate DA estimates of Hispanic undercount are not available. This approximate coverage correction is sufficient for the present purpose of comparing DA and A.C.E. totals to give a rough indication of possible correlation bias. Black Hispanics are a small share of the Black population (about four percent), and less than one percent of the size of the Nonblack population.

- Robinson (2000) explains how alternative Census tabulations corresponding to alternative definitions of the Black and Nonblack race groups can be used in comparing DA to Census results. He considers two extremes for assignment of individuals to the Black and Nonblack groups. Under his Model 1, only those persons who checked only Black for the Census race question are classified as Black. Under his Model 2, persons who checked Black *and* any other race are also classified as Black. In comparing DA and A.C.E. results, however, this affects only the Census group quarters population that is subtracted from the DA totals. In particular, it affects the allocation of the group quarters population to the Black and Nonblack race groups. As can be seen from Table 1 later, this had some effect on the comparison of DA and A.C.E. totals, particularly for Blacks. The effects on the DA sex ratios were negligible, however, thus so were the effects on estimates of correlation bias. Hence, we show only one set of results for sex ratio comparisons and one set for estimating correlation bias.

Comparing DA and A.C.E. estimated population totals

Table 1 compares estimated population totals from DA and A.C.E. for the age-race-sex groups. The table entries show percent coverage differences defined as

$$100 \times (DA - ACE) / DA.$$

If the DA estimates could be taken as "truth," these would represent the percent underestimation or overestimation of the true population by A.C.E. Given the relatively large uncertainties in the DA totals, these are better interpreted simply as percent differences. Table 1.a shows the percent differences in DA and A.C.E. coverage when the census count of group quarters that is subtracted from DA is assigned to the Black and Nonblack race groups using Model 1 described above. Table 1.b shows the percent differences with race assignment of the group quarters population using Model 2. Table 1.c shows for comparison percent coverage differences between 1990 DA estimates and the 1990 PES estimates. Note that in all three tables the entries for age 0-17 are for males and females combined, though listed in the columns labeled "Male."

We draw the following conclusions from these results:

- The most notable finding is that 2000 DA totals are *lower* than A.C.E. totals for many of the age-race-sex groups. The difference is particularly large for Nonblack males and females 18-29 (−8.4 percent and −5.8 percent, respectively.) Comparisons of sex ratios for Nonblacks 18-29 are suspect given these results. Other large differences occur for Black Females 18-29 and for Black children 0-17. All these large negative differences, and even the smaller negative differences, are hard to explain in terms of "correlation bias." They instead suggest some other inconsistency between the DA and A.C.E. estimates.

Table 1.a 2000 A.C.E. Percent Coverage Differences from DA Totals[1]
(Model 1 race assignment of the group quarters population)[2]

Age	Black Male	Black Female	Nonblack Male	Nonblack Female
0-17[3]	**–2.9**	—	–1.6	—
18-29	3.2	**–4.7**	**–8.4**	**–5.8**
30-49	7.6	–.6	–1.6	–2.1
50+	4.2	–.6	.1	–.7

Table 1.b 2000 A.C.E. Percent Coverage Differences from DA Totals[1]
(Model 2 race assignment of the group quarters population)[2]

Age	Black Male	Black Female	Nonblack Male	Nonblack Female
0-17[3]	**–4.0**	—	–1.4	—
18-29	2.0	**–5.8**	**–8.2**	**–5.6**
30-49	7.0	–1.2	–1.5	–2.0
50+	3.9	–.9	.1	–.7

Table 1.c 1990 PES 357 Percent Coverage Differences from DA Totals[1]

Age	Black Male	Black Female	Nonblack Male	Nonblack Female
0-17[3]	1.4	—	–.7	—
18-29	7.8	–.2	–1.8	– 2.1
30-49	7.8	.1	.8	–.9
50+	8.8	.6	2.4	1.2

Notes to Table 1:

1. For 2000, DA estimates revised as of 2/16/01 are used. The DA totals for 2000 and 1990 were adjusted for comparability with the 2000 A.C.E. and 1990 PES results as discussed in the text.

2. Race assignment of the 2000 group quarters population to Black and Nonblack uses one of two "models." Model 1 assigns to Black only those who checked Black and no other race. Model 2 assigns to Black anyone who checked Black, including those who also checked other races.

3. Results shown under "Male" for 0-17 are actually for the total of males and females.

- Comparisons for 1990 show some cases where DA totals are lower than the PES estimates, and the largest such differences again occur for Nonblack Males and Females 18-29. The negative differences, however, are nowhere near as large as those in the 2000 estimates.

- Despite the prevalence of negative entries in Tables 1.a and 1.b, the comparisons for adult Black males show DA totals are significantly larger than the A.C.E. estimates, strongly suggesting correlation bias in the DSEs for these groups. This repeats the general result from 1990, though the 1990 results are more uniform across the age groups. This comparison between years for Black males 50 and over is affected by some recent revisions to the DA estimates that affect this group, as discussed below.

- The differences between entries of Tables 1.a and 1.b are appreciable for Blacks, particularly for the younger age groups, but are quite small for Nonblacks. Even for Blacks, however, these differences do not change any of the above conclusions.

The DA estimates for Blacks 50 and over in 2000 reflect some important revisions for the cohort age 65-75 in 2000 that was 55-65 in 1990. Robinson, et al. (1993) note that DA estimates for this group in 1990 were considered weaker than those for younger and older persons due to the absence of national birth registration data and under registration factors for the years 1925-1935. For 2000, DA estimates for this cohort used Medicare enrollment data. The improved DA estimates for this cohort may account for at least some of the change in coverage differences observed for persons 50 and over between 1990 and 2000.

Comparing DA and A.C.E. estimated sex ratios

Comparisons of DA and A.C.E. sex ratios are given in Table 2. We draw the following conclusions from the results.

- 2000 DA sex ratios for adult Blacks significantly exceed those for A.C.E., strongly suggesting correlation bias in DSEs for adult Black males.

- DA sex ratios for Nonblacks 30-49 and 50 and over only slightly exceed those from A.C.E. suggesting at most small amounts of correlation bias. (To more digits, the Nonblack DA and A.C.E. sex ratios for 30-49 are 1.0020 and .9975, respectively, and those for 50 and over are .8557 and .8494.)

- The DA sex ratio for Nonblacks 18-29 (1.02) is significantly *lower* than that from A.C.E. (1.05). As noted in the discussion of Table 1, this comparison is suspect and probably reflects the underlying inconsistency of the DA and A.C.E. estimates for this group, rather than saying anything about correlation bias. Because of this anomaly, in the next section we do not attempt to estimate correlation bias for Nonblacks 18-29.

- Except for Nonblacks 18-29 and 50 and over, the sex ratio comparisons between DA and

A.C.E. in 2000 are reasonably similar to the corresponding comparisons for 1990. The different treatment of the group quarters populations in the 2000 A.C.E. and 1990 PES could lead to some differences. This may explain the higher sex ratios for Nonblacks 50 and over observed for both the 2000 DA and A.C.E. estimates in comparison to the corresponding 1990 sex ratios.

Table 2.a Sex Ratios from DA[1] and the 2000 A.C.E.

Age	Black[2] A.C.E.	Black[2] DA	Nonblack A.C.E	Nonblack DA
18-29	.84	.90	**1.05**	**1.02**
30-49	.82	.89	1.00	1.00
50+	.72	.76	.85	.86

Table 2.b Sex Ratios from DA and the 1990 PES

Age	Black[2] PES	Black[2] DA	Nonblack PES	Nonblack DA
18-29	.83	.90	1.02	1.02
30-49	.84	.91	.99	1.01
50+	.72	.78	.81	.82

Notes to Table 2:

1. For 2000, DA estimates revised as of 2/16/01 are used. Before computing the DA sex ratios the DA totals were adjusted for comparability with A.C.E. and PES estimates as discussed in the text.

2. Sex ratios for DA in 2000 with race assignment of the group quarters population using Model 1 are the same as those with Model 2 to the accuracy shown in the table.

Estimating Correlation Bias in Adult Male DSEs from the 2000 A.C.E.

Alternative combining models are presented in Bell (1993) and Bell et al. (1996). Statistical refinements to this approach have been developed by Elliott and Little (2000), but their approach would have been impossible to implement in the time allowed, and these refinements are likely to be less important than alternative choices of the combining model. The approach of Elliott and Little (2000) is planned for use later in the total error evaluations.

Correlation bias estimates from the "two-group model"

We begin with results from a particularly simple model that assumes relative correlation bias (percent error) is constant over male post-strata within adult age groups for Blacks and Nonblacks. This model is discussed in Bell (1999). The model can be called the "constant relative bias" model. It has also been called the two-group model because it can be derived by postulating two groups of people within each male post-stratum (say hard-to-count and easy-to-count), and assuming a certain parameter η (that depends on the census and A.C.E. inclusion probabilities and on the proportion of the two groups) is constant across post-strata. The parameter η can also be defined as the ratio of the expected value of the DSE over the true population, and the relative correlation bias is then $\eta - 1$. The parameter η is estimated by the ratio of the A.C.E. sex ratio to the DA sex ratio, data for which are given in Table 2.

Table 3 gives estimation results for the two-group model for the 2000 A.C.E. and 1990 PES, in terms of the relative correlation bias estimates expressed as percents. The results show the following:

- Two-group model estimates of correlation bias for Blacks 18-29 and 30-49 in 2000 are very similar to those in 1990, both implying around eight percent underestimation by the DSEs.

- The estimate of correlation bias for Blacks 50 and over in 2000 (–4.7 percent) is smaller in size than in 1990 (–8.2 percent). This may be partly due to the revisions to the DA estimates for the cohort of Black males who were 65-75 in 2000, as discussed earlier in connection with the results in Table 1.

- Estimates of correlation bias for Nonblack Males 30-49 and 50 and over are negative and quite small in magnitude. Similar results are obtained for 1990, though the estimates are slightly larger in magnitude.

- The estimate for Nonblacks 18-29, +2.5 percent, is hard to explain as "correlation bias." Possible causes of such a result are that males in this group who were more likely than others to be missed in the census were more likely than others to be found in the A.C.E., or that females in this group exhibit a larger amount of conventional correlation bias than do males.

Neither of these explanations makes much sense. A better interpretation of the result for this group is simply that because of the inconsistency between DA and A.C.E. results, we cannot estimate correlation bias for Nonblacks 18-29.

- Given this conclusion, we decided to proceed by assuming no correlation bias for Nonblack males 18-29 (as well as for Nonblack females 18-29) on the grounds that
  - estimates of correlation bias for older Nonblack males are small in 2000, and
  - the estimate of correlation bias for Nonblack males 18-29 in 1990 was small.

Table 3.a Relative Correlation Bias Estimates[1] for 2000 A.C.E.
(results from two-group model[2], expressed as percents)

Age	Black	Nonblack
18-29	–7.4	**2.5**
30-49	–8.1	–.5
50+	–4.7	–.7

Table 3.b Relative Correlation Bias Estimates[1] for 1990 PES
(results from two-group model[2], expressed as percents)

Age	Black	Nonblack
18-29	–8.0	–.3
30-49	–7.7	–1.6
50+	–8.2	–1.2

Notes to Table 3:

1. These estimates use the DA sex ratios from Table 2.

2. For the two-group model the estimate of relative correlation bias, expressed as a percent, is

$$100\left(\frac{\text{A.C.E. sex ratio}}{\text{DA sex ratio}} - 1\right).$$

Correlation bias estimates from alternative models

We estimated correlation bias for the male post-strata using the alternative models listed in Table 4. The first model is the two-group model already discussed. Models 2-5 were presented in Bell (1993) and are discussed also in Elliott and Little (2000). Model 6 is discussed by Das Gupta in Appendix B of Bell et al. (1996). All the models assume no correlation bias for females, that DA sex ratios are accurate, and that some quantity related to correlation bias (model parameter) is constant over male post-strata within age-race groups. Estimation of the model parameter and of correlation bias for the male post-strata then follows these steps:

1. Aggregate DSEs for females to the national level within age-race groups (Black vs. Nonblack).

2. Multiply the female totals by DA sex ratios to get control totals for males.

3. Determine the model parameter so that the resulting estimates for male post-strata (which are unbiased assuming the model is true) aggregate to the control totals from step 2.

4. Difference between these estimates and the usual DSEs estimate correlation bias for adult male post-strata. Relative (percent) differences estimate relative (percent) correlation bias.

If N is the true population in a post-stratum then the (arithmetic) correlation bias in the DSE is E(DSE) − N and the relative correlation bias is [E(DSE) − N] / N.

Estimates of the parameters of the alternative models are given in Table 5 for the adult age groups for Black and Nonblack. Results are given for both the 2000 A.C.E. and 1990 PES. Results for 2000 for Nonblacks 18-29 are included only to allow a point to be made below about the inconsistency between DA and A.C.E. for this group. The estimated models for this group are not used later to produce correlation bias estimates by post-strata.

Estimates of the two-group model parameter η are simply one plus the relative correlation bias estimates given in Table 3. Estimation results for models 2-5 for the 1990 PES (357 post-strata) were given in Bell (1993). For the FOR, FRR, and MF22 models, values of the parameters exceeding 1 indicate negative correlation bias resulting in underestimation by DSEs. For the two-group, BRE, and Das Gupta models values of the parameters less than 1 indicate negative correlation bias resulting in DSE underestimation. Actually, for the MF22, BRE, and Das Gupta this is not quite true, since the DSE assuming independence is not a special case of these models. For these three models under the conditions cited it is more appropriate to say there is a tendency for the usual DSEs to underestimate, though this need not be uniformly true across post-strata.

Table 4. Alternative Models[1] for Estimating Correlation Bias in DSEs Using DA Sex Ratios

Model	Assumption
1. Two-group or fixed relative bias (FRB)	Correlation bias[2] proportional to population of post-stratum, implying relative correlation bias[2] constant over post-strata.
2. Fixed odds ratio (FOR)	Correlation bias in post-stratum proportional to persons missed by both the Census and A.C.E., the (2,2) cell.
3. Fixed relative risk (FRR)	Correlation bias in post-stratum proportional to persons missed by the Census.
4. Fixed ratio of male to female p_{22} (MF22)	Ratio of male to female probabilities for (2,2) cell constant over post-strata.
5. Generalized behavioral response estimator (BRE)	Pr[in ACE | not in Census] / Pr[in Census] constant over post-strata.
6. Prithwis Das Gupta's model (PDG)	Pr[in ACE or Census for males] / Pr[in ACE or Census for females] constant over post-strata.

Notes to Table 4:

1. The models derive from alternative assumptions that some quantity (parameter) is constant across all male post-strata within an age-race (Nonblack versus Black) group. Model 1 was discussed earlier and is described in Bell (1999, Attachment A). Models 2-5 are taken from Bell (1993) and are also discussed by Elliott and Little (2000). Model 6 is discussed by Das Gupta in Appendix B of Bell, et al. (1996). Consult these references for mathematical details.

2. If N is the true population then (arithmetic) correlation bias is E(DSE)−N and relative correlation bias is [E(DSE)−N] / N.

Table 5. Parameter Estimates for Alternative Models[1] to Estimate Correlation Bias
Results for 2000 A.C.E.

	Age	2-group (η)	FOR (θ)	FRR (γ)	MF22 (ρ)	BRE (λ)	Das Gupta
Blacks	18-29	.93	3.34	1.44	5.44	.76	.91
	30-49	.92	4.86	1.60	7.96	.69	.91
	50+	.95	5.17	1.49	8.69	.71	.95
Nonblacks	18-29	1.02	0.00	.82	0.00	1.24	1.02
	30-49	1.00	1.81	1.06	3.71	1.00	.99
	50+	.99	3.02	1.12	3.97	.92	.99

Results for 1990 PES

	Age	2-group (η)	FOR (θ)	FRR (γ)	MF22 (ρ)	BRE (λ)	Das Gupta
Blacks	18-29	.92	2.87	1.43	3.50	.76	.91
	30-49	.92	2.94	1.47	5.30	.72	.90
	50+	.92	6.81	1.91	8.50	.58	.91
Nonblacks	18-29	1.00	1.19	1.02	1.77	1.01	.99
	30-49	.98	3.64	1.23	6.17	.86	.98
	50+	.99	4.05	1.26	4.26	.81	.99

Note to Table 5.

1. The alternative models are those listed in Table 4. For the FOR, FRR, and MF22 models values of the parameters exceeding 1 indicate negative correlation bias resulting in underestimation by DSEs. For the 2-group, BRE, and Das Gupta models values of the parameters less than 1 indicate negative correlation bias resulting in DSE underestimation. Actually, for the MF22, BRE, and Das Gupta this is not quite true, since the DSE assuming independence is not a special case of these models. For these models under the conditions cited it is more appropriate to say there is a tendency for the usual DSEs to underestimate, though this need not be uniformly true across post-strata.

The results in Table 5 suggest the following conclusions:

- From Table 2 (with more digits than shown for some of the entries) DA sex ratios exceed those from the A.C.E. and PES estimates suggesting negative correlation bias except for Nonblacks 18-29 in 2000. Corresponding estimates of the various model parameters are greater or less than 1 as predicted from the above discussion except for Nonblacks 18-29 in 2000 and one other minor exception for the BRE model (Nonblacks 18-29 in 1990).

- Parameter estimates for Nonblacks 18-29 in 2000 make little sense. In fact, the estimates of the FOR and MF22 model parameters are set at their boundary values of 0, and even these value did not allow these models to reproduce the control totals noted in estimation step 3 above. With their parameters set to zero these models simply add up the observed cells of the post-stratum 2×2 tables. For Nonblack males 18-29 the aggregation of these results, which allow for no persons missed by both the census and A.C.E., still exceed the control totals. This strongly suggests that, whatever is the reason for the discrepancy between the DA and A.C.E. results for Nonblacks 18-29, it is not due to correlation bias, even positive correlation bias, for males.

- In several cases parameter estimates for Blacks 18-29 and 30-49 in 2000 are similar to those in 1990, though there are also exceptions to this. Parameter estimates for Blacks 50 and over differ between 1990 and 2000, as would be expected since the relation between the DA and A.C.E. sex ratios for this group differ between the two years. Apart from the meaningless results for Nonblacks 18-29 in 2000, parameter estimates for Nonblacks 30-49 and 50 and over are not very similar between 1990 and 2000.

As suggested earlier, estimates of correlation bias for individual adult male post-strata can be obtained as $DSE - N_{alt}$, where N_{alt} is the population estimate from the alternative model, which is unbiased under the model assumptions. Estimates of relative correlation bias are then $[DSE - N_{alt}] / N_{alt}$. Table 6 (at the end of this memo) gives these estimates of relative correlation bias obtained from the alternative models for 2000 for all the adult male post-strata except those for Nonblacks 18-29. Post-stratum groups that were collapsed over age were synthetically uncollapsed proportional to the individual post-stratum census counts as described in Bell (1993).

For the two-group model the entries in Table 6 are constant over post-strata for a given age-race group because this model assumes that relative correlation bias is constant in this way. For the other models the estimates of relative correlation bias vary significantly over post-strata, showing that the model assumed for estimating correlation bias by post-strata has a significant impact on the results. As noted in Bell (1993), all the alternative models fit the data equally well (all are saturated models), and so the data give no information about whether one or the other of the models should be preferred. For this reason, when using post-stratum correlation bias estimates in such things as loss function analyses (see Mulry and Spencer 1993), it is useful to obtain results using correlation bias estimates from several of these alternative models, to check if the results of interest depend on the model used.

The Wachter-Freedman Criticism

Wachter and Freedman (1999) criticized Census Bureau efforts to estimate correlation bias in the 1990 PES DSEs (Bell 1991, Mulry and Spencer 1993) for failing to allow for other estimated biases when estimating correlation bias. In principle, this criticism is valid. However, Spencer (2000) noted that if the other biases are not very different between males and females, then, when using DA sex ratios, the effects of these other biases on estimates of correlation bias for males will be small. Spencer's argument as stated applies only to the aggregate (national level) estimate of correlation bias, however, not to correlation bias estimates for individual post-strata.

For the two-group model Spencer's argument extends to show that estimates of correlation bias for male post-strata are unaffected by other (multiplicative) biases that are the same for males and females. For the other models used we cannot assert that their estimates of correlation bias for individual post-strata will be unaffected by other biases. However, all these other models produce post-stratum population estimates that aggregate to the same quantities at the national level as the two-group model. Therefore, for all the models we can at least say that these aggregates are unaffected by other biases that are the same for males and females (since this holds for the two-group model). Although, in the presence of other biases, the alternative models won't produce estimates of correlation bias for individual post-strata that are unbiased under the model assumptions, the estimates nonetheless have some validity since they satisfy the aggregation constraint. Because of this, while their correlation bias estimates don't conform perfectly to the model assumptions, they still provide alternative allocations of the aggregate estimates of correlation bias across the male post-strata within age-race groups. Therefore, these alternative estimates should provide reasonable results for investigating the sensitivity of various analyses (e.g., of the total error model and loss function analyses) to alternative assumptions about correlation bias.

Evaluations to estimate other biases in the A.C.E. DSEs will be carried out as part of the total error model analysis, but those results will not be available until late 2001. With these results it may be possible to adjust the underlying data for the DSEs for other biases to address the Wachter-Freedman criticism at the post-stratum level for any of the alternative models.

For the present results we have assumed the "other biases" are approximately the same for males and females. Bench (2000) provides some tabulations of the 1990 PES error evaluations that are relevant to this issue. Her results do not directly address the issue as directly as may be desired since they involve taking estimates of the other biases for "evaluation post-strata" (Mulry and Spencer 1993), disaggregating down to individual post-strata according to various assumptions, and then aggregating the results back up for various tables. The effect of this disaggregation and then aggregation on the results is not immediately clear. Nonetheless, putting this aside, some relevant information is provided in the last table of her Appendix E, which shows the difference between the original PES sex ratios and sex ratios obtained from her aggregates of the DSEs modified for the other biases. Results are given for the three adult age groups, but are not broken out by race. The results show what appear to be minor differences between the original and modified sex ratios, suggesting little impact of the other biases on the A.C.E. sex ratios. This suggests that, in the aggregate, the other biases in the 1990 PES estimates may not have been very different for males and females. Conclusions are tentative, and direct tabulations of results from the evaluation post-strata to the national level broken out by Black and Nonblack race groups would be more desirable. Of course, later results from the 2000 evaluations will be even more relevant.

References

Bell, William, “1990 Post Enumeration Survey Evaluation Project P13: Use of Alternative Dual System Estimators to Measure Correlation Bias,” 1990 Coverage Studies and Evaluation Memorandum Series, U.S. Bureau of the Census, 1991.

Bell, William, “Using Information from Demographic Analysis in Post-Enumeration Survey Estimation,” Journal of the American Statistical Association, 88, pp. 1106-1118, 1993.

Bell, William, “Accuracy and Coverage Evaluation Survey: Combining Logistic Regression DSEs (Target Model) with 1990 Demographic Analysis Results,” DSSD Census 2000 Procedures and Operations Memorandum Series Q-12, September 22, 1999.

Bell, William, Campbell Gibson, Prithwis Das Gupta, Gregory Spencer, J. Gregory Robinson, Mary Mulry, Elizabeth Vacca, Robert Fay, and Charlene Leggieri, “Report of the Working Group on the Use of Demographic Analysis in Census 2000,” U.S. Bureau of the Census, 1996.

Bench, Katie, “Computations of Proportion of Bias in the 1990 DSE and DSE Sex Ratios – Revision 1,” Planning, Research, and Evaluation Division TXE/2010 Memorandum Series CM-TE-S-02R1, December 12, 2000.

Elliott, Michael and Roderick Little, “A Bayesian Approach to Combining Information from a Census, a Coverage Measurement Survey, and Demographic Analysis,” Journal of the American Statistical Association, 95, pp. 351-362, 2000.

Griffin, Richard, “Accuracy and Coverage Evaluation Survey: Post-stratification Research Methodology,” DSSD Census 2000 Procedures and Operations Memorandum Series Q-5, September 23, 1999.

Griffin, Richard, “Accuracy and Coverage Evaluation Survey: Definition and Explanation of Correlation and Related Heterogeneity Bias,” DSSD Census 2000 Procedures and Operations Memorandum Series Q-35, September 29, 2000.

Mulry, Mary, “1990 Post Enumeration Survey Evaluation Project P16: Total Error in PES Estimates for Evaluation Post-strata,” 1990 Coverage Studies and Evaluation Memorandum Series R-6, U.S. Bureau of the Census, July 11, 1991.

Mulry, Mary and Bruce Spencer, “Total Error in PES Estimates of Population (with discussion),” Journal of the American Statistical Association, 86, pp. 839-863, 1991.

Mulry, Mary and Bruce Spencer, “Accuracy of the 1990 Census and Undercount Adjustments,” Journal of the American Statistical Association, 88, pp. 1080-1091, 1993.

Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta, and Karen Woodrow, “Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis (with discussion),” Journal of the American Statistical Association, 88, pp. 1061-1079, 1993.

Robinson, J. Gregory, “Accuracy and Coverage Evaluation Survey: Demographic Analysis Results,” DSSD Census 2000 Procedures and Operations Memorandum Series B-4*, February 28, 2001.

Schindler, Eric, “Accuracy and Coverage Evaluation Survey: Post-stratification Preliminary Research Results,” DSSD Census 2000 Procedures and Operations Memorandum Series Q-23, January 12, 2000.

Spencer, Bruce, “Final Report on Correlation Bias Methodology,” report prepared under contract number 50-YABC-7-66020 for the U.S. Bureau of the Census, Abt Associates, Inc. and Spencer Statistics, Inc., April 20, 2000.

Wachter, Kenneth and David Freedman, “The Fifth Cell: Correlation Bias in U.S. Census Adjustment,” Technical Report Number 570, Department of Statistics, University of California, Berkeley, 1999.

Zaslavsky, Alan and Glenn Wolfgang, “Triple System Modeling of Census, Post-Enumeration Survey, and Administrative List Data,” Journal of Business and Economic Statistics, 11, pp. 279-288, 1993.

Table 6. Relative Correlation Bias Estimates for Adult Male Post-strata from Alternative Models
Table 6.a Blacks 18-29

post-stratum	two-group	FOR	FRR	MF22	BRE	Das Gupta
41	-7.37	-.69	-4.75	-2.49	-5.56	-9.01
42	-7.37	-8.85	-6.28	-16.80	-2.18	-8.42
43	-7.37	-7.06	-5.87	-4.09	-2.87	-7.01
44	-7.37	-1.92	-8.78	-13.75	-16.98	-10.50
45	-7.37	-11.83	-8.31	-11.07	-4.86	-6.16
46	-7.37	-4.74	-9.71	-.27	-18.67	-7.25
47	-7.37	-8.61	-8.06	-6.25	-7.48	-6.75
48	-7.37	-4.46	-7.28	.41	-8.96	-7.24

Table 6.b Blacks 30-49

post-stratum	two-group	FOR	FRR	MF22	BRE	Das Gupta
41	-8.10	-2.94	-5.90	-3.58	-5.74	-8.87
42	-8.10	-12.28	-9.02	-16.96	-6.95	-8.04
43	-8.10	-5.48	-5.19	-3.33	-3.30	-8.26
44	-8.10	.00	-7.27	.00	-9.95	-9.09
45	-8.10	-13.56	-9.70	-15.38	-7.88	-7.48
46	-8.10	-7.89	-11.39	-4.16	-17.74	-7.77
47	-8.10	-8.95	-9.19	-3.12	-9.65	-7.37
48	-8.10	-7.14	-9.62	-.34	-12.08	-7.52

Table 6.c Blacks 50 and over

post-stratum	two-group	FOR	FRR	MF22	BRE	Das Gupta
41	-4.74	-2.03	-3.14	-1.60	-2.72	-5.02
42	-4.74	-4.67	-4.47	-3.18	-3.72	-4.62
43	-4.74	-5.20	-3.90	-4.63	-2.53	-4.65
44	-4.74	.00	-3.02	.00	-3.33	-5.27
45	-4.74	-9.19	-6.48	-13.99	-5.70	-4.71
46	-4.74	-4.39	-8.25	1.11	-13.60	-4.21
47	-4.74	-8.84	-7.18	-8.19	-7.71	-4.16
48	-4.74	-7.66	-8.06	-.55	-11.10	-3.62

Table 6.d Nonblacks 30-49

post-stratum	two-group	FOR	FRR	MF22	BRE	Das Gupta
1	-.45	.00	-.28	.00	-.28	-.75
2	-.45	.00	-.19	.00	-.13	-.75
3	-.45	-.09	-.27	.11	-.14	-.64
4	-.45	.00	-.28	.00	-.27	-.75
5	-.45	-.35	-.30	.34	.17	-.34
6	-.45	-.84	-.31	-2.01	.82	-.52
7	-.45	-1.63	-.45	-3.20	1.64	-.09
8	-.45	-.43	-.46	.53	-.19	-.22
9	-.45	-.11	-.21	-.39	.00	-.75
10	-.45	.00	-.16	.00	-.08	-.75
11	-.45	.00	-.27	.00	-.25	-.75
12	-.45	.00	-.25	.00	-.21	-.75
13	-.45	.00	-.15	.00	-.07	-.75
14	-.45	-.70	-.34	-1.65	.57	-.55
15	-.45	-.77	-.41	-2.35	.49	-.67
16	-.45	-1.52	-.68	.84	.52	.89
17	-.45	-.46	-.28	-.36	.35	-.43
18	-.45	-.06	-.21	.07	-.07	-.67
19	-.45	-.02	-.28	.03	-.25	-.72
20	-.45	-.56	-.28	.42	.50	-.13
21	-.45	.00	-.35	.00	-.44	-.75
22	-.45	.00	-.60	.00	-1.52	-.75
23	-.45	-.50	-.49	-.24	-.20	-.35
24	-.45	-1.53	-.57	-3.73	1.08	-.33
25	-.45	-.54	-.33	-1.62	.35	-.69
26	-.45	-.28	-.18	-.40	.27	-.60
27	-.45	-.69	-.36	-.42	.52	-.23
28	-.45	.00	-.61	.00	-1.56	-.75

Table 6.d Nonblacks 30-49 (continued)

post-stratum	two-group	FOR	FRR	MF22	BRE	Das Gupta
29	-.45	.00	-.30	.00	-.31	-.75
30	-.45	-.72	-.33	.90	.61	.15
31	-.45	-.38	-.49	.47	-.37	-.28
32	-.45	-2.27	-.73	-1.21	1.51	1.08
33	-.45	-.10	-.63	-.30	-1.48	-.74
34	-.45	-1.26	-.95	-1.85	-2.05	-.08
35	-.45	-.54	-.58	-.74	-.53	-.46
36	-.45	-1.58	-.89	.33	-.88	.81
37	-.45	-.67	-.64	-1.19	-.68	-.46
38	-.45	.00	-.90	.00	-3.80	-.75
39	-.45	-.51	-.67	-.60	-1.08	-.44
40	-.45	-.21	-.89	.26	-3.41	-.49
49	-.45	-.29	-.41	.15	-.22	-.44
50	-.45	-1.05	-.63	.15	.02	.26
51	-.45	-.94	-.47	-1.81	.55	-.37
52	-.45	-1.64	-.45	-4.43	1.66	-.40
53	-.45	-1.55	-.81	.19	-.33	.74
54	-.45	.00	-1.11	.00	-6.51	-.75
55	-.45	-2.12	-1.00	-3.57	-1.01	.29
56	-.45	-1.96	-1.55	2.55	-11.10	1.78
57	-.45	-.61	-.92	-.42	-2.92	-.31
58	-.45	-.49	-.77	-3.95	-1.85	-1.36
59	-.45	-.23	-.41	-.75	-.31	-.74
60	-.45	-.82	-.75	-2.83	-1.13	-.76
61	-.45	-3.32	-.80	-9.53	2.94	-.08
62	-.45	-2.18	-.69	-5.59	1.54	-.19
63	-.45	-1.15	-.43	-8.35	.99	-1.93
64	-.45	-2.37	-.80	-4.73	1.23	.24

Table 6.e Nonblacks 50 and over

post-stratum	two-group	FOR	FRR	MF22	BRE	Das Gupta
1	-.74	.00	-.49	.00	-.54	-.83
2	-.74	-.38	-.30	-.21	-.06	-.75
3	-.74	-.56	-.56	-.17	-.33	-.67
4	-.74	.00	-.44	-.02	-.47	-.84
5	-.74	-1.14	-1.02	-1.51	-.95	-.79
6	-.74	-2.06	-.70	-2.23	.36	-.61
7	-.74	-2.36	-.67	-2.78	.60	-.63
8	-.74	-1.06	-1.02	.53	-1.00	-.31
9	-.74	-.91	-.43	-1.12	.07	-.78
10	-.74	.00	-.28	-.06	-.26	-.85
11	-.74	-.23	-.60	.11	-.60	-.72
12	-.74	.00	-.45	.00	-.48	-.83
13	-.74	.00	-.33	.00	-.31	-.83
14	-.74	-.73	-.53	-.98	-.17	-.81
15	-.74	-2.63	-.85	-1.61	.43	-.23
16	-.74	-3.09	-.94	-2.42	.55	-.25
17	-.74	-.84	-.39	-1.90	.09	-1.00
18	-.74	-.23	-.40	-.46	-.28	-.86
19	-.74	.00	-.50	.00	-.56	-.83
20	-.74	-.26	-.39	-.50	-.25	-.86
21	-.74	.00	-.77	.00	-1.06	-.83
22	-.74	.00	-.78	.00	-1.09	-.83
23	-.74	.00	-.83	.00	-1.19	-.83
24	-.74	-1.09	-.70	-.31	-.24	-.50
25	-.74	-.24	-.99	.12	-1.45	-.71
26	-.74	-1.01	-.53	-1.28	-.01	-.78
27	-.74	-.88	-.94	-1.32	-.91	-.84
28	-.74	.00	-1.16	.00	-2.11	-.83

Table 6.e Nonblacks 50 and over (continued)

post-stratum	two-group	FOR	FRR	MF22	BRE	Das Gupta
29	-.74	-.58	-1.21	.29	-1.89	-.55
30	-.74	-1.28	-.73	-.49	-.20	-.48
31	-.74	-1.04	-.77	-.87	-.41	-.66
32	-.74	-1.76	-1.29	-3.15	-1.32	-.96
33	-.74	-.09	-1.14	-1.75	-1.97	-1.24
34	-.74	-3.18	-1.43	-4.69	-.78	-.78
35	-.74	-.26	-1.02	-1.18	-1.52	-1.03
36	-.74	-4.66	-2.09	-3.80	-2.68	.06
37	-.74	-1.36	-1.11	-.34	-1.04	-.41
38	-.74	-.22	-1.77	-.24	-4.48	-.81
39	-.74	.00	-1.42	.00	-3.03	-.83
40	-.74	-1.62	-2.03	-.15	-4.86	-.26
49	-.74	-.94	-.65	-1.04	-.25	-.75
50	-.74	-3.59	-1.11	-2.25	.48	.00
51	-.74	-.55	-.67	-.89	-.52	-.85
52	-.74	-1.63	-.90	-2.53	-.34	-.85
53	-.74	-3.14	-1.50	-1.34	-1.10	.04
54	-.74	.00	-2.30	.00	-8.01	-.83
55	-.74	-.35	-2.14	-1.73	-6.58	-1.14
56	-.74	-4.77	-3.41	2.55	-15.57	1.69
57	-.74	-1.51	-2.03	-.50	-4.93	-.39
58	-.74	-1.22	-1.72	-4.28	-3.45	-1.45
59	-.74	-1.31	-.85	-.86	-.43	-.56
60	-.74	-2.50	-1.37	-3.71	-1.09	-.81
61	-.74	-7.72	-1.94	-6.46	.78	.76
62	-.74	-5.64	-1.37	-20.66	1.24	-3.73
63	-.74	-1.65	-.81	.83	-.11	.00
64	-.74	-1.97	-1.05	1.00	-.48	.16